

Stochastic Modeling in Systems Biology

Jinzhi Lei*

Zhou Pei-Yuan Center for Applied Mathematics
Tsinghua University, Beijing, 100084, China

January 20, 2013

Abstract

Many cellular behaviors are regulated by gene regulation networks, kinetics of which is one of the main subjects in the study of systems biology. Because of the low number molecules in these reacting systems, stochastic effects are significant. In recent years, stochasticity in modeling the kinetics of gene regulation networks have been drawing the attention of many researchers. This paper is a self contained review trying to provide an overview of stochastic modeling. I will introduce the derivation of the main equations in modeling the biochemical systems with intrinsic noise (chemical master equation, Fokker-Plan equation, reaction rate equation, chemical Langevin equation), and will discuss the relations between these formulations. The mathematical formulations for systems with fluctuations in kinetic parameters are also discussed. Finally, I will introduce the exact stochastic simulation algorithm and the approximate explicit tau-leaping method for making numerical simulations.

1 Introduction

Systems biology is an interdisciplinary science of discovering, modeling, understanding and ultimately engineering at the molecular level the dynamic relationships between the biological molecules that define living organisms¹. This field is increasingly hot in recent decades as modeling molecular systems is not only fascinating but also possible in the post-genomics science. At the molecular level, many cellular behaviors are regulated by genetic regulation networks in which the stochasticity is significant. To describe these stochastic chemical kinetics, stochastic modeling is highlighted recently [20, 29, 51].

Chemical dynamics have been widely accepted to study chemical kinetics of reacting systems with large molecule populations, typically in the order of 10^{23} . In these systems, the kinetics are nearly deterministic and can be described by a set of ordinary differential equations—the *reaction rate equations*, or partial differential equations if spatial movements are taken into account. Nevertheless, in intracellular molecule kinetics, stochasticity is significant because the numbers

*Email: jzlei@mail.tsinghua.edu.cn

¹According to Dr. Leroy Hood, the first president of the Institute of Systems Biology, Seattle. http://www.systemsbiology.org/Systems_Biology_in_Depth

of each molecule species are very low. For instance, only one gene, either active or inactive, is involved in most activities of gene expressions. In an individual bacteria, there are less than 20 transcriptions of mRNAs from a single gene [22]. Typical molecule numbers of the same protein specie in a cell are usually no more than a few thousands. Thus, fluctuations in protein activities are significant due to the low-number effect. Reaction rate equations fail to describe these fluctuations. In this review, I will introduce main equations for the stochastic modeling of such systems, and discuss numerical methods used in stochastic simulations.

This paper will first briefly review assumptions and notations for describing a biochemical system, followed by two examples of biological processes. Next, I will introduce several theoretical formulations for modeling biochemical system kinetics with merely intrinsic noise, followed by the mathematical formulations for the situation in which kinetic parameters are random. Finally, some stochastic simulation methods supported by the previous theories are introduced. The paper will be concluded with a summarization of the theoretical structure of stochastic modeling.

2 Stochasticity in biological processes

It is no doubt that biological processes are essentially random [14, 20, 33, 35, 36, 42, 44, 48, 51]. Both cellular behavior and the cellular environment are stochastic. Phenotypes vary across isogenic populations and in individual cell over time [51]. Gene expression is a fundamentally stochastic process, with randomness in transcription and translation leading to cell-to-cell variations in mRNA and protein levels [10, 29, 40]. Noise propagation in gene networks has important consequences for cellular functions, being beneficial in some contexts and harmful in others [41, 42].

Essentially, kinetics of biological molecules in living cells are consequences of chemically reacting systems. In this section, we first review basic assumptions and descriptions of general chemical systems, and then give two examples to demonstrate their applications.

2.1 Chemical systems

Consider a system of well-stirred mixture of $N \geq 1$ molecular species $\{S_1, \dots, S_N\}$, inside some fixed volume Ω and at constant temperature, through $M \geq 1$ reaction channels $\{R_1, \dots, R_M\}$. We can specify the dynamical state of this system by $\mathbf{X}(t) = (X_1(t), \dots, X_N(t))$, where

$$X_i(t) = \text{the number of } S_i \text{ molecule in the system at time } t, \quad (t = 1, \dots, N). \quad (1)$$

We will described the evolution of $\mathbf{X}(t)$ from some given initial state $\mathbf{X}(t_0) = \mathbf{x}_0$. It is obvious that $\mathbf{X}(t)$ is a stochastic process, because the time at which a particular reaction occurs is random. Therefore, instead of tracking a single pathway, our goal is to study the evolution of the statistical properties of system states.

In this review paper, we always assume that each reaction, once occur, completes instantaneously. This is to be distinguished with the systems involve

reactions with delay [49]. Further, we assume that the system is well stirred such that at any moment, each reactions occur with equal probability at any position. Under these assumptions, each reaction channel R_j associates with a *propensity function* a_j and a *state-change vector* $\mathbf{v}_j = (v_{j1}, \dots, v_{jN})$, which are defined such that

$$a_j(\mathbf{x})dt = \begin{aligned} &\text{the probability, given } \mathbf{X}(t) = \mathbf{x}, \text{ that one reaction } R_j \\ &\text{will occur somewhere inside } \Omega \text{ in the next infinitesimal} \\ &\text{time interval } [t, t + dt], \quad (j = 1, \dots, M). \end{aligned} \quad (2)$$

and

$$v_{ji} = \begin{aligned} &\text{the change in the number of } S_i \text{ molecule produced} \\ &\text{by one } R_j \text{ reaction,} \quad (j = 1, \dots, M; i = 1, \dots, N). \end{aligned} \quad (3)$$

The propensity function and the state-change vector together specify the reaction channel R_j . Therefore, the equations given below to describe the evolution of a biochemical system are derived from the propensity functions and stage-change vectors connected to the M reaction channels.

The state-change vector of a reaction channel is easy to be obtained by counting the numbers of each molecule species that are consumed and produced in one reaction. For instance, if R_1 were the reaction $S_1 + 2S_2 \rightarrow 2S_3$, then $\mathbf{v}_1 = (-1, -2, 2, \dots)$.

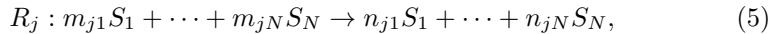
Exact descriptions of propensity functions associate with the *ad hoc* stochasticity of deterministic chemical kinetics [39], and have solid microphysical basis. In general, the function a_j have the mathematical form [18]

$$a_j(\mathbf{x}) = c_j h_j(\mathbf{x}). \quad (4)$$

Here c_j is the *specific probability rate constant* for the channel R_j , which is defined such that $c_j dt$ is the probability that a randomly chosen combination of R_j reactant molecules will react accordingly in the next infinitesimal time interval dt . This probability $c_j dt$ equals the multiple of two parts, the probability of a randomly chosen combination of R_j reactant molecules will collide in the next dt , and the probability that a colliding reactant molecules will actually react according to R_j . The first probability depends on the average relative speeds (which in turn depends on the temperature), the collision sections of a reactant molecules, and the system volume Ω . The second probability depends on the chemical energy barrier $\Delta\mu$ of the reaction R_j (Figure 1), and usually associate with temperature through a Boltzmann factor $e^{-\Delta\mu/k_B T}$, where k_B is the Boltzmann constant.

The function $h_j(\mathbf{x})$ in (4) measures the number of distinct combinations of R_j reactant molecules available in the state \mathbf{x} . It can be easily obtained from the reaction R_j . For example, in the above reaction R_1 , we would have $h_j(\mathbf{x}) = x_1 x_2 (x_2 - 1)/2$, which give the number of combinations to select one S_1 molecule from x_1 of them, and two S_2 molecules from x_2 of them. More examples are given below.

In general, for a chemical reaction



we would have

$$v_{ji} = n_{ji} - m_{ji}, \quad a_j(\mathbf{x}) \propto \prod_{k=1}^N \frac{x_k!}{m_{jk}!(x_k - m_{jk})!}.$$

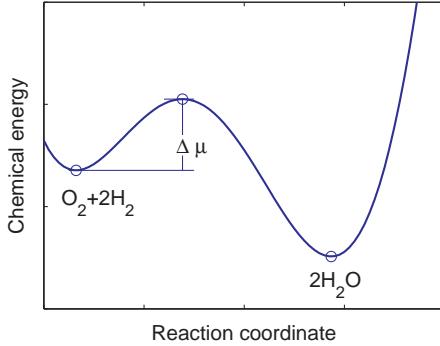


Figure 1: Schematic of chemical energy for the reaction $O_2 + 2H_2 \rightarrow 2H_2O$.

If for any k and j , have $x_k \gg m_{jk}$, then approximately

$$a_j(\mathbf{x}) = c_j \prod_{k=1}^N x_k^{m_{jk}}.$$

The *reaction rate constant* c_j can only be obtained from experiments, and usually depends on the system volume Ω and the temperature.

In real systems, most reaction channels are either *monomolecular* or *bimolecular* reaction. For a monomolecular reaction, the reaction rate constant c_j is independent of the system volume Ω . For a bimolecular reaction, the rate constant c_j is inversely proportional Ω . *Trimolecular* reactions do not physically occur in dilute solutions with appreciable frequency. One can consider a trimolecular reaction as the combined result of two bimolecular reactions, and involved an additional short-lived species. For such an “effective trimolecular” reaction, the approximate c_j is proportional to Ω^{-2} [18].

2.2 Examples in gene regulations

2.2.1 Gene expression

Gene expression is a basic biological process. Reactions in gene expression include promoter activity and inactivity, *transcription*, *translation*, and decaying of mRNA and proteins. Typical steps in gene expression are illustrated in Figure 2 (also refer [32]). Note that transcription is a process of mRNA synthesis as specified by the gene, in which only the information is read out and the gene is not consumed. Similarly, proteins are translated from a mRNA sequence according to the *genetic code*, and the mRNA is not consumed in this process.

In many gene regulations, the transition between active and inactive promoter states are regulated by a proteins (activator or repressor). In the case of activation, the activator bind to the inactive promoter to enhance the gene expression. The reaction channels R_1 and R_2 become

$$R_1 : X_5 + X_1 \rightarrow X_2, \quad R_2 : X_2 \rightarrow X_1 + X_5,$$

where X_5 stands for the number of the activator. The corresponding propensity functions and state-change vectors are $a_1(\mathbf{X}) = c_1 X_1 X_5, a_2(\mathbf{X}) = c_2 X_2, \mathbf{v}_1 =$

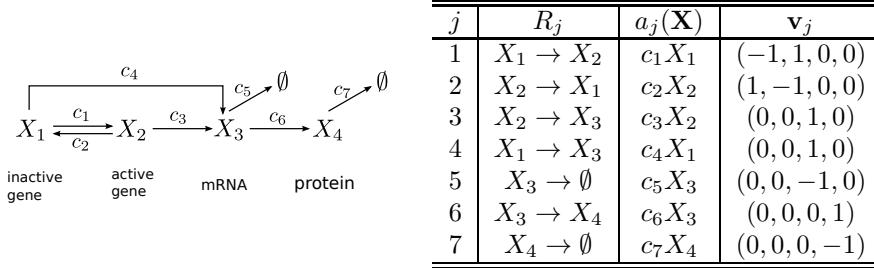


Figure 2: (From ref. [32]) A model of gene expression. Each step represents the biochemical reactions which associate with transition between promoter states, production and decaying of mRNAs and proteins (here $c_3 > c_4$).

$(-1, 1, 0, 0, -1)$, and $\mathbf{v}_2 = (1, -1, 0, 0, 1)$. Similarly, in the case of repressor, we should have

$$R_1 : X_1 \rightarrow X_2 + X_5, \quad R_2 : X_5 + X_2 \rightarrow X_1,$$

and $a_1(\mathbf{X}) = c_1 X_1$, $a_2(\mathbf{X}) = c_2 X_2 X_5$, $\mathbf{v}_1 = (-1, 1, 0, 0, 1)$, and $\mathbf{v}_2 = (1, -1, 0, 0, -1)$.

2.2.2 Circadian oscillator

Figure 3 shows a simple model of circadian oscillator based on a common positive and negative control elements found experimentally [57]. In this model, two genes, an activator A and a repressor R , are transcribed into mRNA and subsequently translated into protein. The activator A binds to both A and R promoters to increase their transcription rates. The protein R binds to and sequester the protein A , and therefore acts as a repressor. Figure 3 shows the propensity functions, and stage-change vectors of this model (c_i in the Figure gives the rate constant of the reaction channel R_i). Note that in the reaction channel R_{16} , the complex breaks in to R because of the degradation of A . Thus R_{16} is not the reversion process of R_{15} .

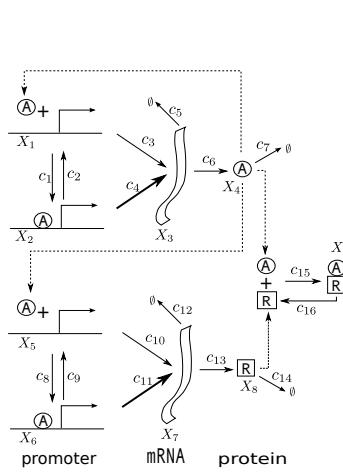
3 Mathematical formulations–intrinsic noise

First, we assume that the propensity functions are time independent, i.e., the reaction rates c_j in (4) are constants. In this situation, the fluctuations in the system are inherent to the system of interest (*intrinsic noise*). The opposite case is *extrinsic noise*, which arises from variability in factors that are considered to be external. Mathematical formulations for extrinsic noise will be discussed in next section.

3.1 Chemical master equation

From the propensity function given in previous, the state vector $\mathbf{X}(t)$ is a jump-type Markov process on the non-negative N -dimensional integer lattice. In following analysis of such a system, we will focus on the conditional *probability function*

$$P(\mathbf{x}, t | \mathbf{x}_0, t_0) = \text{Prob}\{\mathbf{X}(t) = \mathbf{x}, \text{ given that } \mathbf{X}(t_0) = \mathbf{x}_0\}. \quad (6)$$



j	$a_j(\mathbf{X})$	\mathbf{v}_j
1	$c_1 X_1 X_4$	$(-1, 1, 0, -1, 0, 0, 0, 0, 0)$
2	$c_2 X_2$	$(1, -1, 0, 1, 0, 0, 0, 0, 0)$
3	$c_3 X_1$	$(0, 0, 1, 0, 0, 0, 0, 0, 0)$
4	$c_4 X_2$	$(0, 0, 1, 0, 0, 0, 0, 0, 0)$
5	$c_5 X_3$	$(0, 0, -1, 0, 0, 0, 0, 0, 0)$
6	$c_6 X_3$	$(0, 0, 0, 1, 0, 0, 0, 0, 0)$
7	$c_7 X_4$	$(0, 0, 0, -1, 0, 0, 0, 0, 0)$
8	$c_8 X_5 X_4$	$(0, 0, 0, -1, -1, 1, 0, 0, 0)$
9	$c_9 X_6$	$(0, 0, 0, 1, 1, -1, 0, 0, 0)$
10	$c_{10} X_5$	$(0, 0, 0, 0, 0, 1, 0, 0, 0)$
11	$c_{11} X_6$	$(0, 0, 0, 0, 0, 0, 1, 0, 0)$
12	$c_{12} X_7$	$(0, 0, 0, 0, 0, -1, 0, 0)$
13	$c_{13} X_7$	$(0, 0, 0, 0, 0, 0, 0, 1, 0)$
14	$c_{14} X_8$	$(0, 0, 0, 0, 0, 0, 0, -1, 0)$
15	$c_{15} X_4 X_8$	$(0, 0, 0, -1, 0, 0, 0, -1, 1)$
16	$c_{16} X_9$	$(0, 0, 0, 0, 0, 0, 0, 1, -1)$

Figure 3: (From ref. [57]) A biochemical network of the circadian oscillator model.

Hereinafter, we use an upper case letter to denote a random variable, and the corresponding lower case letter for a possible value of that random variable.

Through the probability function, the *average* or *expectation value* of any quantity $f(\mathbf{X}|\mathbf{x}_0, t_0)$ defined on the system is given by

$$\langle f(\mathbf{X}|\mathbf{x}_0, t_0) \rangle = \sum_{\mathbf{x}} f(\mathbf{x}) P(\mathbf{x}, t|\mathbf{x}_0, t_0). \quad (7)$$

Physically, in an ensemble of identical systems starting from the same initial state $\mathbf{X}(t_0) = \mathbf{x}_0$, the function $P(\mathbf{x}, t|\mathbf{x}_0, t_0)$ gives the fraction of subsystems with state $\mathbf{X}(t) = \mathbf{x}$.

To derive a time evolution for the probability function $P(\mathbf{x}, t|\mathbf{x}_0, t_0)$, we take a time increment dt and consider the variation between the probability of $\mathbf{X}(t) = \mathbf{x}$ and of $\mathbf{X}(t+dt) = \mathbf{x}$, given that $\mathbf{X}(t_0) = \mathbf{x}_0$. This variation is

$$\begin{aligned} P(\mathbf{x}, t+dt|\mathbf{x}_0, t_0) - P(\mathbf{x}, t|\mathbf{x}_0, t_0) &= \text{Increasing of the probability in } dt \\ &\quad - \\ &\quad \text{Decreasing of the probability in } dt \end{aligned} \quad (8)$$

We take dt so small such that the probability of having two or more reactions in dt is negligible compared to the probability of having only one reaction. Then increasing of the probability in dt occurs when a system with state $\mathbf{X}(t) = \mathbf{x} - \mathbf{v}_j$ reacts according to R_j in $(t, t+dt)$, the probability of which is $a_j(\mathbf{x} - \mathbf{v}_j)dt$. Thus,

$$\text{Increasing of the probability in } dt = \sum_{j=1}^M P(\mathbf{x} - \mathbf{v}_j, t|\mathbf{x}_0, t_0) a_j(\mathbf{x} - \mathbf{v}_j) dt. \quad (9)$$

Similarly, when a system with state $\mathbf{X}(t) = \mathbf{x}$ reacts according to any reaction

channel R_j in $(t, t + dt)$, the probability $P(\mathbf{x}, t | \mathbf{x}_0, t_0)$ will decrease. Thus,

$$\text{Decreasing of the probability in } dt = \sum_{j=1}^M P(\mathbf{x}, t | \mathbf{x}_0, t_0) a_j(\mathbf{x}) dt. \quad (10)$$

Substituting (9) and (10) into (8), we obtain

$$\begin{aligned} P(\mathbf{x}, t + dt | \mathbf{x}_0, t_0) - P(\mathbf{x}, t | \mathbf{x}_0, t_0) &= \sum_{j=1}^M P(\mathbf{x} - \mathbf{v}_j, t | \mathbf{x}_0, t_0) a_j(\mathbf{x} - \mathbf{v}_j) dt \\ &\quad - \sum_{j=1}^M P(\mathbf{x}, t | \mathbf{x}_0, t_0) a_j(\mathbf{x}) dt, \end{aligned}$$

which yields, with the limit $dt \rightarrow 0$, the *chemical master equation* [18, 56]:

$$\frac{\partial}{\partial t} P(\mathbf{x}, t | \mathbf{x}_0, t_0) = \sum_{j=1}^M [P(\mathbf{x} - \mathbf{v}_j, t | \mathbf{x}_0, t_0) a_j(\mathbf{x} - \mathbf{v}_j) - P(\mathbf{x}, t | \mathbf{x}_0, t_0) a_j(\mathbf{x})]. \quad (11)$$

The equation (11) is an exact consequence from the reaction channels characterized by propensity functions and stage-change vectors. If one can solve (11) for P , we should be able to find out everything about the process $\mathbf{X}(t)$. However, such an exact solution of (11) can rarely be obtained (refer [27, 50] for the examples of solving the chemical master equation analytically).

In fact, the chemical master equation (11) is a set of linear differential equations with constant coefficients. The difficult on solving the equation (11) comes from the high dimension, which equals the total number of possible states of the system under study. For example, for a system with 100 molecules species, each has two possible states ($X_i(t) = 0$ or 1), the system has totally 2^{100} possible states, and therefore the equation (11) constants 2^{100} equations! Even solving such a huge system numerically is a big challenge.

3.2 Fokker-Plank equation

If in a biochemical system, all components of $\mathbf{X}(t)$ are very large compared to 1, we can regard the components of $\mathbf{X}(t)$ as real numbers. We assume further that the functions $f_j(\mathbf{x}) \equiv a_j(\mathbf{x})P(\mathbf{x}, t | \mathbf{x}_0, t_0)$ are analytic in the variable \mathbf{x} . With these two assumptions, we can use Taylor's expansion to write

$$f_j(\mathbf{x} - \mathbf{v}_j) = f_j(\mathbf{x}) + \sum_{|\mathbf{m}| \geq 1} \prod_{i=1}^N \frac{(-1)^{m_i}}{m_i!} \left(\frac{\partial}{\partial x_i} \right)^{m_i} (v_{ji}^{m_i} f_j(\mathbf{x})). \quad (12)$$

Here $\mathbf{m} = (m_1, \dots, m_N) \in \mathbb{Z}^N$, $|\mathbf{m}| = m_1 + \dots + m_N$. Substituting (12) into (11), we immediately obtain the *chemical Kramer-Moyal equation* (CKME) [18]

$$\frac{\partial}{\partial t} P(\mathbf{x}, t) = \sum_{|\mathbf{m}| \geq 1} \prod_{i=1}^N \frac{(-1)^{m_i}}{m_i!} \left(\frac{\partial}{\partial x_i} \right)^{m_i} (A^{m_1, \dots, m_N}(\mathbf{x}) P(\mathbf{x}, t)), \quad (13)$$

where

$$A^{m_1, \dots, m_N}(\mathbf{x}) = \sum_{j=1}^M v_{j1}^{m_1} \cdots v_{jN}^{m_N} a_j(\mathbf{x}).$$

Hereinafter, we omit the initial condition (\mathbf{x}_0, t_0) . If all functions $f_j(\mathbf{x} - \mathbf{v}_j)$ are smooth, the equation (13) would equivalent to (11). Therefore, the chemical Kramer-Moyal equation is a “semi-rigorous” consequence of the chemical master equation (11).

If we truncate the right hand side at $|\mathbf{m}| = 2$, we obtain the *chemical Fokker-Plank equation* (CFPE)

$$\frac{\partial}{\partial t} P(\mathbf{x}, t) = - \sum_{i=1}^N \frac{\partial}{\partial x_i} A_i(\mathbf{x}) P(\mathbf{x}, t) + \frac{1}{2} \sum_{1 \leq i, k \leq N} \frac{\partial^2}{\partial x_i \partial x_k} B_{ik}(\mathbf{x}) P(\mathbf{x}, t). \quad (14)$$

where

$$A_i(\mathbf{x}) = \sum_{j=1}^M v_{ji} a_j(\mathbf{x}), \quad B_{ik}(\mathbf{x}) = \sum_{j=1}^M v_{ji} v_{jk} a_j(\mathbf{x}). \quad (15)$$

We will re-obtain this chemical Fokker-Plank equation below from the chemical Langevin equation.

3.3 Reaction rate equation

If we multiply the chemical master equation (11) through by x_i , and sum over all \mathbf{X} , we obtain the following *chemical ensemble average equation* (CEAE)

$$\frac{d\langle X_i \rangle}{dt} = \sum_{j=1}^M v_{ji} \langle a_j(\mathbf{X}) \rangle \quad (i = 1, 2, \dots, N). \quad (16)$$

Equation (16) is an exact consequence of (11). Nevertheless, (16) is not a close equation system unless all reaction channels are monomolecular. When all reaction channels are monomolecular, the propensity functions $a_j(\mathbf{X})$ are linear, and therefore

$$\langle a_j(\mathbf{X}) \rangle = a_j(\langle \mathbf{X} \rangle) \quad (j = 1, \dots, N). \quad (17)$$

Thus, the populations evolve deterministically according to a set of ordinary differential equations

$$\frac{dx_i}{dt} = \sum_{j=1}^M v_{ji} a_j(\mathbf{x}) \quad (i = 1, \dots, N), \quad (18)$$

where the components of $\mathbf{x}(t)$ are now considered as real variables.

In usual, the mean populations do not evolve according to (18) when there are higher order reactions. However, (18) is sometimes heuristic if we assuming that the fluctuations are not important and (17) holds approximately. Consequently, the ordinary differential equations (18) also hold under this assumption. The equation (18) is referred as the macroscopic *reaction rate equation*(RRE), or *chemical rate equation* in some literatures.

The reaction rate equation (18) is often written in terms of the species concentrations

$$Z_i(t) \equiv X_i(t)/\Omega \quad (i = 1, \dots, N).$$

The “concentrations” form of the reaction rate equation has form

$$\frac{dz_i}{dt} = \sum_{j=1}^M v_{ji} \tilde{a}_j(\mathbf{z}) \quad (i = 1, \dots, N), \quad (19)$$

where

$$\tilde{a}_j(\mathbf{z}) = a_j(\Omega \mathbf{z})/\Omega.$$

While we examine the Ω dependence in the propensity functions for monomolecular, bimolecular, and trimolecular reactions, we would find that the function \tilde{a}_j is functionally identical to a_j except that the rate constants c_j has been replaced by the *reaction rate constants* k_j .

The reaction rate equations are most commonly used in modeling biochemical reacting systems. However, this equation fails to describe the stochastic effects which can be very important in biological processes. We will introduce the *chemical Langevin equation* below that has a facility to describe the stochasticity, and easy to be studied, at least numerically.

3.4 Chemical Langevin equation

The chemical Langevin equation was derived to yield an approximate time-evolution equation of the Langevin type. The derivation of the equation, given by Gillespie, is based on the chemical master equation and two explicit dynamical conditions as detailed below. Most of the following refer to Gillespie's original paper [18].

Suppose that the state of a system at current time t is $\mathbf{X}(t) = \mathbf{x}$. Let $K_j(\mathbf{x}, \tau)$ ($\tau > 0$) be the number of R_j reactions that occur in the subsequent time interval $[t, t + \tau]$. Since each of these reactions will change the S_i population by v_{ji} , the number of S_i molecule in the system at time $t + \tau$ will be

$$X_i(t + \tau) = x_i + \sum_{j=1}^M K_j(\mathbf{x}, \tau) v_{ji}, \quad (i = 1, \dots, N). \quad (20)$$

We note that $K_j(\mathbf{x}, \tau)$ is a *random variable*, and therefore $X_i(t + \tau)$ is random.

We will obtain an approximation of $K_j(\mathbf{x}, \tau)$ below by imposing the following conditions:

Condition (i) Require τ to be *small* enough such that the change in the state during $[t, t + \tau]$ will be so slight that none of the propensity functions changes its value "appreciably".

Condition (ii) Require τ to be *large* enough that the expected number of occurrences of each reaction channel R_j in $[t, t + \tau]$ be much larger than 1.

From condition (i), the propensity functions satisfy

$$a_j(\mathbf{X}(t')) \approx a_j(\mathbf{x}), \quad \forall t' \in [t, t + \tau], \forall j \in [1, M]. \quad (21)$$

Thus, the probability of the reaction R_j to occur in any infinitesimal interval $d\tau$ within $[t, t + \tau]$ is $a_j(\mathbf{x})d\tau$. Thus, $K_j(\mathbf{x}, \tau)$, the occurrence of "events" of reaction R_j in the time interval $[t, t + \tau]$, will be a statistically independent *Poisson* random variable, and is denoted by $\mathcal{P}_j(a_j(\mathbf{x}), \tau)$. So (20) can be approximated by

$$X_i(t + \tau) = x_i + \sum_{j=1}^M v_{ji} \mathcal{P}_j(a_j(\mathbf{x}), \tau), \quad (i = 1, \dots, N) \quad (22)$$

according to condition (i).

The mean and variance of $\mathcal{P}(a, \tau)$ are

$$\langle \mathcal{P}(a, \tau) \rangle = \text{var}\{\mathcal{P}(a, \tau)\} = a\tau. \quad (23)$$

Thus, condition (ii) means

$$a_j(\mathbf{x})\tau \gg 1, \quad \forall j \in [1, M]. \quad (24)$$

The inequality (24) allows us to approximate each Poisson random variable by a *normal* random variable with the same mean and variance. This leads to further approximation

$$X_i(t + \tau) = x_i + \sum_{j=1}^M v_{ji} \mathcal{N}_j(a_j(\mathbf{x})\tau, a_j(\mathbf{x})\tau), \quad (i = 1, \dots, N) \quad (25)$$

where $\mathcal{N}_j(m, \sigma^2)$ denotes the normal random variable with mean m and variance σ^2 . Here the M random variables are independent to each other. Notice that in the above approximation, we have converted the molecular population X_i from discretely changing integers to continuously changing real variables.

We note the linear combination theorem for normal random variables,

$$\mathcal{N}(m, \sigma^2) = m + \sigma \mathcal{N}(0, 1), \quad (26)$$

the equation (25) can be rewritten as

$$X_i(t + \tau) = x_i + \sum_{j=1}^M v_{ji} a_j(\mathbf{x})\tau + \sum_{j=1}^M v_{ji} \sqrt{a_j(\mathbf{x})\tau} \mathcal{N}_j(0, 1) \quad (i = 1, \dots, N). \quad (27)$$

Now, we are ready to obtain Langevin type equations by making some purely notational changes. First, we denote the time interval τ by dt , and write

$$dX_i = X_i(t + dt) - X_i(t).$$

Next, introduce M temporally uncorrelated, independent random process $W_j(t)$, satisfying

$$dW_j(t) = W_j(t + dt) - W_j(t) = \mathcal{N}_j(0, 1) \sqrt{dt} \quad (j = 1, \dots, N). \quad (28)$$

It is easy to verify that the processes $W_j(t)$ have *stationary independent increments with mean 0*, i.e.,

$$\langle dW_j(t) \rangle = 0, \quad \langle dW_i(t) dW_j(t') \rangle = \delta_{ij} \delta(t - t') dt, \quad \forall 1 \leq i, j \leq M, \forall t, t'. \quad (29)$$

Thus, each W_j is a *Wiener process* (or referred to as *Brownian motion*) [56]. Finally, recalling that \mathbf{x} stands for $\mathbf{X}(t)$, the equation (25) becomes a *Langevin equation* (or *stochastic differential equation*)

$$dX_i = \sum_{j=1}^M v_{ji} a_j(\mathbf{X}) dt + \sum_{j=1}^M v_{ji} \sqrt{a_j(\mathbf{X})} dW_j \quad (i = 1, \dots, N). \quad (30)$$

The stochastic differential equation (30) is the desired *chemical Langevin equation* (CLE). The solution of (30) with initial condition $\mathbf{X}(0) = \mathbf{X}_0$ is a stochastic process $\mathbf{X}(t)$ satisfying

$$\mathbf{X}(t) = \mathbf{X}_0 + \sum_{j=1}^M \int_0^t v_{ji} a_j(\mathbf{X}(s)) ds + \sum_{j=1}^M \int_0^t v_{ji} \sqrt{a_j(\mathbf{X}(s))} dW_j(s). \quad (31)$$

Here *Itô integral* is used as the stochastic fluctuations are *intrinsic* [56].

In the above, it is obvious that the conditions (i) and (ii) are contradict to each other. It may be very well happen that both conditions cannot be satisfied simultaneously. In this case, the derivation of the chemical Langevin equation may fail. But there are many practical circumstances in which the two conditions can be simultaneously satisfied. As the inequality (24) implies that $a_j(\mathbf{x})$ is large when τ is small enough as required by condition (i). This is possible when the system has large molecular populations for each molecule species since $a_j(\mathbf{x})$ is typically proportional to one or more components of \mathbf{x} .

Even when the conditions (i) and (ii) cannot be satisfied simultaneously, the chemical Langevin equation (30) is still useful. As we will discuss below.

3.5 Discussions about the chemical Langevin equation

In many intracellular biochemical systems such as gene expressions and genetic networks, the molecule populations are small, and the reactions are usually slow. In these systems, the derivation of the chemical Langevin equation may fail. Nevertheless, the equation (30) is still useful for such systems as it can provide reasonable descriptions for the statistical properties of the kinetic processes. The reasons are given below.

3.5.1 Time-evolution of the mean

If we take average to both sides of (30), noticing

$$\langle \sqrt{a_j(\mathbf{X})} dW_j \rangle = 0 \quad (j = 1, \dots, M)$$

according to the Itô interpretation, we have

$$\frac{d\langle X_i \rangle}{dt} = \sum_{j=1}^M v_{ji} \langle a_j(\mathbf{X}) \rangle \quad (i = 1, \dots, N). \quad (32)$$

This gives the same form of chemical ensemble average equation (16) as we have obtained from the chemical master equation.

3.5.2 Time-evolution of the correlations

The correlations between the molecule numbers are defined as

$$\sigma_{ij}(t) = \langle X_i(t) X_j(t) \rangle - \langle X_i(t) \rangle \langle X_j(t) \rangle, \quad (1 \leq i, j \leq N). \quad (33)$$

Multiply the chemical master equation (11) through by $x_i x_j$ and sum over all \mathbf{x} , we obtain

$$\frac{d\langle X_i X_j \rangle}{dt} = \sum_{k=1}^M [\langle X_i v_{kj} a_k(\mathbf{X}) \rangle + \langle X_j v_{ki} a_k(\mathbf{X}) \rangle + \langle v_{ki} v_{kj} a_k(\mathbf{X}) \rangle]. \quad (34)$$

Furthermore, (16) gives

$$\frac{d\langle X_i \rangle \langle X_j \rangle}{dt} = \sum_{k=1}^M [\langle v_{ki} a_k(\mathbf{X}) \rangle \langle X_j \rangle + \langle v_{kj} a_k(\mathbf{X}) \rangle \langle X_i \rangle]. \quad (35)$$

Thus, recalling (15), we have the *chemical ensemble correlations equation* (CECE)

$$\frac{d\sigma_{ij}}{dt} = \langle A_i(\mathbf{X})(X_j - \langle X_j \rangle) \rangle + \langle (X_i - \langle X_i \rangle) A_j(\mathbf{X}) \rangle + \langle B_{ij}(\mathbf{X}) \rangle. \quad (36)$$

The equation (36) is *exact* from the chemical master equation.

Now, starting from the chemical Langevin equation (30) and applying the Itô formula [38], we have

$$\begin{aligned} d(X_i X_j) &= X_i(dX_j) + X_j(dX_i) + (dX_i)(dX_j) \\ &= \sum_{k=1}^M [X_j v_{ki} a_k(\mathbf{X}) + X_i v_{kj} a_k(\mathbf{X}) X_i + v_{ki} v_{kj} a_k(\mathbf{X})] dt \\ &\quad + \sum_{k=1}^M [X_i v_{kj} \sqrt{a_k(\mathbf{X})} + X_j v_{ki} \sqrt{a_k(\mathbf{X})}] dW_k + o(dt) \end{aligned}$$

Taking the average to both sides, we re-obtain (34), which yields the same form of chemical ensemble correlations equation (36).

At states of near equilibrium, if the random fluctuations are not important, we have approximately

$$A_i(\mathbf{X}) \approx A_i(\langle \mathbf{X} \rangle) + \sum_{l=1}^N \frac{\partial A_i(\langle \mathbf{X} \rangle)}{\partial X_l} (X_l - \langle X_l \rangle), \quad \langle B_{ij}(\mathbf{X}) \rangle \approx B_{ij}(\langle \mathbf{X} \rangle).$$

Substituting above approximations into (36), and defining the matrices

$$\sigma = (\sigma_{ij}), \quad A = \left(\frac{\partial A_i(\langle \mathbf{X} \rangle)}{\partial X_l} \right), \quad B = (B_{ij}(\langle \mathbf{X} \rangle)), \quad (37)$$

we have

$$\frac{d\sigma}{dt} = (A\sigma + \sigma A^T) + B. \quad (38)$$

Equation (38) gives the linear approximation of the time-evolution of the correlation functions near equilibrium. This is an example of the well known *fluctuation-dissipation theorem* [56].

3.5.3 Fokker-Plank equation

Consider the chemical Langevin equation (30), the solution $\mathbf{X}(t)$ satisfies

$$X_i(t + \Delta t) = X_i(t) + \sum_{k=1}^M v_{ki} a_j(\mathbf{X}) \Delta t + \sum_{k=1}^M v_{ki} \sqrt{a_k(\mathbf{X})} \Delta W_k(t), \quad (i = 1, \dots, N) \quad (39)$$

where

$$\Delta W_k(t) = W_k(t + \Delta t) - W_k(t).$$

Thus, assuming $\mathbf{X}(t) = \mathbf{x}$, the displacement $\Delta\mathbf{x} = \mathbf{X}(t + \Delta t) - \mathbf{X}(t)$ satisfies

$$\langle \Delta x_i \rangle = \sum_{k=1}^M v_{ki} a_k(\mathbf{x}) \Delta t$$

and

$$\langle \Delta x_i \Delta x_j \rangle = \sum_{k=1}^M v_{ki} v_{kj} a_k(\mathbf{x}) \Delta t.$$

Let $P(\mathbf{x}, t)$ to be the conditional probability density function as previous, then $P(\mathbf{x}, t)$ satisfies

$$\begin{aligned} P(\mathbf{x}, t + dt) - P(\mathbf{x}, t) &= \int_{\Delta\mathbf{x} \in \mathbb{R}^N} P(\mathbf{x} - \Delta\mathbf{x}, t) W(\Delta\mathbf{x}, dt; \mathbf{x} - \Delta\mathbf{x}, t) d\Delta\mathbf{x} \\ &\quad - \int_{\Delta\mathbf{x} \in \mathbb{R}^N} P(\mathbf{x}, t) W(\Delta\mathbf{x}, dt; \mathbf{x}, t) d\Delta\mathbf{x}, \end{aligned}$$

where $W(\Delta\mathbf{x}, dt; \mathbf{x}, t)$ is the transition probability from $\mathbf{X}(t) = \mathbf{x}$ to $\mathbf{X}(t + dt) = \mathbf{x} + \Delta\mathbf{x}$. Applying Taylor expansion to the function

$$h(\mathbf{x} - \Delta\mathbf{x}) \equiv P(\mathbf{x} - \Delta\mathbf{x}, t) W(\Delta\mathbf{x}, dt; \mathbf{x} - \Delta\mathbf{x}, t),$$

noticing

$$\int_{\Delta\mathbf{x} \in \mathbb{R}^N} \Delta x_i W(\Delta\mathbf{x}, dt; \mathbf{x}, t) d\Delta\mathbf{x} = \langle \Delta x_i \rangle,$$

and

$$\int_{\Delta\mathbf{x} \in \mathbb{R}^N} \Delta x_i \Delta x_j W(\Delta\mathbf{x}, dt; \mathbf{x}, t) d\Delta\mathbf{x} = \langle \Delta x_i \Delta x_j \rangle,$$

we have

$$\begin{aligned} P(\mathbf{x}, t + dt) - P(\mathbf{x}, t) &= - \sum_{i=1}^N \frac{\partial}{\partial x_i} [P(\mathbf{x}, t) \langle \Delta x_i \rangle] \\ &\quad + \frac{1}{2} \sum_{1 \leq i, j \leq N} \frac{\partial^2}{\partial x_i \partial x_j} [P(\mathbf{x}, t) \langle \Delta x_i \Delta x_j \rangle]. \end{aligned}$$

Thus, letting $dt \rightarrow 0$ (here $\Delta t = dt$), we re-obtain the chemical Fokker-Plank equation (14) as previous.

Thus, the chemical Langevin equation and the chemical master equation yield the same form of chemical Fokker-Plank equation.

3.5.4 Remarks

In comparing the chemical Langevin equation and chemical master equation, we should note the following:

1. In this section, we have shown that the chemical Langevin equation and chemical master equation yield the same forms of chemical ensemble average equation (16), chemical ensemble correlation equation (36), and chemical Fokker-Plank equation (14).

2. Despite the same forms of equations (16) and (36), since the averages are taken with respect to probability functions $P(\mathbf{x}, t)$, which are obtained from either the solutions of the chemical master equation or the chemical Langevin equation, the two equations (CLE and CME) do not always yield the same dynamics of ensemble average and correlation. These are possible when all reaction channels are monomolecular, in which cases the two equations (16) and (36) are, as we have discussed before, close.
3. The chemical Langevin equation and chemical master equation yield the same form of chemical Fokker-Plan equation, which is a close equation. Thus, up to the second order approximation, the two equations give the same time-evolutions of the probability function $P(\mathbf{x}, t)$.

4 Mathematical formulations–fluctuation in kinetic parameters

In the previous discussion, we assume that the reaction rates c_j in (4) are constants. This is only a rough approximation of the real world, in which cell environments are stochastic, and therefore the reaction rates are random. Here, we will discuss the mathematical formulations for situations in which there are fluctuations in kinetic parameters.

4.1 Reaction rate as a random process

When the reaction rate c_j is random, the previous propensity function $a_j(\mathbf{x})$ in (4) should be rewritten as

$$a_j(\mathbf{x}, t) = c_j(t)h_j(\mathbf{x}). \quad (40)$$

Here $c_j(t)$, as previous, is the specific probability reaction rate for channel R_j at time t . This rate is usually random, depending on the fluctuating environment whose explicit time dependence is not known. A reasonable approximation is given below.

In previous discussions, we have obtained $c_j \propto e^{-\Delta\mu_j/k_B T}$. If there are noise perturbations to the energy barrier, we can replace $\Delta\mu_j$ by $\Delta\bar{\mu}_j - k_B T\eta_j(t)$, where $\eta_j(t)$ is a stochastic process. Consequently, we can replace the reaction rate $c_j(t)$ by

$$c_j(t) = \bar{c}_j e^{\eta_j(t)} / \langle e^{\eta_j(t)} \rangle, \quad (41)$$

where \bar{c}_j is a constant, measures the mean of the reaction rate.

In many cases, the noise perturbation $\eta(t)$ (here we omit the subscript j) can be described by an *Ornstein-Uhlenbeck process*, which is given by a solution of following stochastic differential equation

$$d\eta = -(\eta/\tau)dt + (\sigma/\tau)dW. \quad (42)$$

Here W is a Wiener process, τ and σ are positive constants, measuring the autocorrelation time and variance, respectively. It is easy to verify that $\eta(t)$ is normally distributed and has an exponentially decaying stationary autocorrelation function [17, 54, 55, 56]

$$\langle \eta(t)\eta(t') \rangle = \frac{\sigma^2}{2\tau} e^{-|t-t'|/\tau}. \quad (43)$$

The Ornstein-Uhlenbeck process is an example of *color noise*. When $\tau \rightarrow 0$, $\eta(t)$ approaches to *Gaussian white noise*.

With $\eta(t)$ an Ornstein-Uhlenbeck process, the stationary distribution of the reaction rate $c_j(t)$ is then *log-normal*. Log-normal distribution have been seen in many applications [34]. For instance, log-normal rather than normal distribution have been measured for gene expression rates [45, 52].

When $\eta(t)$ is an Ornstein-Uhlenbeck process, we have [8]

$$\langle e^{\eta(t)} \rangle = e^{\sigma^2/(4\tau)}. \quad (44)$$

Therefore, the reaction rate (41) can be rewritten as

$$c_j(t) = \bar{c}_j e^{\eta_j(t) - \sigma_j^2/(4\tau_j)}. \quad (45)$$

Thus, the propensity function (40) now becomes

$$a_j(\mathbf{x}, t) = \bar{c}_j e^{\eta_j(t) - \sigma_j^2/(4\tau_j)} h_j(\mathbf{x}), \quad (46)$$

with $\eta_j(t)$ satisfies an equation of form (42).

Next, we will introduce generalizations of the reaction rate equation and the chemical Langevin equation to describe reaction systems with fluctuations in reaction rates. We note that in this situation, the state variable \mathbf{X} alone is not enough to describe the state of a system, and the state of environment has to be taken into account as well. Thus, to generalize the chemical master equation or the Fokker-Plank equation, we should replace the previous probability function with $P(\mathbf{x}, \mathbf{c}, t | \mathbf{x}_0, \mathbf{c}_0, t)$, where $\mathbf{c} = \{c_1, \dots, c_M\}$. This consideration is omitted here.

4.2 Reaction rate equation

Substituting the propensity functions of form (46) and (42) into the reaction rate equation (18), we obtain the following stochastic differential equations

$$dX_i = \left(\sum_{j=1}^M v_{ji} \bar{c}_j e^{\eta_j - \sigma_j^2/(4\tau_j)} h_j(\mathbf{X}) \right) dt, \quad (i = 1, \dots, N) \quad (47)$$

$$d\eta_j = -(\eta_j/\tau_j)dt + (\sigma_j/\tau_j)dW_j, \quad (j = 1, \dots, M) \quad (48)$$

Here $\bar{c}_j, \tau_j, \sigma_j$ are positive constants. The equations (47)-(48) generalize the reaction rate equation to describe the reaction system with fluctuations in the kinetic parameters.

4.3 Chemical Langevin equation

Similar to the above strategy, substituting the propensity functions of form (46) and (42) into the chemical Langevin equation (30), we obtain the following

stochastic differential equations

$$dX_i = \left(\sum_{j=1}^M v_{ji} \bar{c}_j e^{\eta_j - \sigma_j^2/(4\tau_j)} h_j(\mathbf{X}) \right) dt \quad (49)$$

$$+ \sum_{j=1}^M v_{ji} \left(\bar{c}_j e^{-\eta_j - \sigma_j^2/(4\tau_j)} h_j(\mathbf{X}) \right)^{1/2} dW_j, \quad (i = 1, \dots, N)$$

$$d\eta_j = -(\eta_j/\tau_j)dt + (\sigma_j/\tau_j)dW'_j, \quad (j = 1, \dots, M) \quad (50)$$

Here W'_j are independent Wiener processes, and $\bar{c}_j, \tau_j, \sigma_j$, as previous, are positive constants. This set of generalized chemical Langevin equations describes the stochasticity of chemical systems with both intrinsic noise and fluctuations in kinetic parameters.

5 Stochastic simulations

We have introduced several equations for modeling systems of biochemical reactions. Here, we will introduce stochastic simulation methods that intend to mimic a random process $\mathbf{X}(t)$ of a system evolution. Once we have enough sampling pathways of the random process, we are able to calculate the probability density function $P(\mathbf{x}, t)$ and other statistical behaviors including the mean trajectory and correlations.

5.1 Stochastic simulation algorithm

Assume the system in state \mathbf{x} at time t . We define the *next-reaction density function* $p(\tau, j|\mathbf{x}, t)$ as

$$p(\tau, j|\mathbf{x}, t)d\tau = \text{probability that, given } \mathbf{X}(t) = \mathbf{x}, \text{ the next reaction in } \Omega \text{ will occur in the infinitesimal time interval } [t + \tau, t + \tau + d\tau), \text{ and will be an } R_j \text{ reaction.} \quad (51)$$

An elementary probability argument based on the propensity function (2) yields [15, 16]

$$p(\tau, j|\mathbf{x}, t) = a_j(\mathbf{x}) \exp(-a_0(\mathbf{x})\tau), \quad (0 \leq \tau < \infty, j = 1, \dots, M) \quad (52)$$

where

$$a_0(\mathbf{x}) = \sum_{k=1}^M a_k(\mathbf{x}).$$

This formula provides the basis for the *stochastic simulation algorithm*(SSA) (also known as the *Gillespie algorithm*).

Following direct method is perhaps the simplest to generate a pair of numbers (τ, j) in accordance with the probability (52) [16]: First generate two random numbers r_1 and r_2 of uniform distribution in the unit interval, and then take

$$\tau = (1/a_0(\mathbf{x})) \ln(1/r_1) \quad (53)$$

$$j = \text{the smallest integer satisfying } \sum_{j'=1}^j a_{j'}(\mathbf{x}) > r_2 a_0(\mathbf{x}). \quad (54)$$

Below are main steps in the stochastic simulation algorithm (for details, refer [16])

1. **Initialization:** Let $\mathbf{X} = \mathbf{x}_0$, and $t = t_0$.
2. **Monte Carlo step:** Generate a pair of random numbers (τ, j) according to the probability density function (52) (replace \mathbf{x} by \mathbf{X}).
3. **Update:** Increase the time by τ , and replace the molecule count by $\mathbf{X} + \mathbf{v}_j$.
4. **Iterate:** Go back to Step 2 unless the simulation time has been exceeded.

The stochastic simulation algorithm consists with the chemical master equation in the sense that (11) and (52) are exact consequence of the propensity function (2). This algorithm numerically simulation the time evolution of a given chemical system, and gives a sample trajectory of the real system.

5.2 Tau-leaping algorithm

In the derivation of the chemical Langevin equation, we can always select τ small enough such that the the following *leap condition* is satisfied:

Leap condition: During $[t, t + \tau)$, no propensity function is likely to change its value by a significant amount.

Consequently, the previous arguments indicate that we can approximately leap the system with a time τ by taking

$$\mathbf{X}(t + \tau) \approx \mathbf{x} + \sum_{j=1}^M \mathcal{P}_j(a_j(\mathbf{x}), \tau) \mathbf{v}_j, \quad (55)$$

where $\mathbf{x} = \mathbf{X}(t)$.

The equation (55) is the basic of the *tau-leaping algorithm* [19]: Starting from the current state \mathbf{x} , we first choose a value τ that satisfies the leap condition. Next, we generate for each j a random number k_j according to Poisson distribution with mean $a_j(\mathbf{x})\tau$. Finally, we update the state from \mathbf{x} to $\mathbf{x} + \sum_j k_j \mathbf{v}_j$, and increase the time by τ .

There are two practical issues need to be resolved in order to effectively apply the tau-leaping algorithm: First, how can we estimate the largest value of τ that satisfies the leap condition? Second, how can we ensure that the generated k_j values do not result in negative populations?

The original estimation of the largest value of τ was given by Gillespie and is sketched below [19]. If τ satisfies the leap condition, from (55), the average state changes over time τ is

$$\bar{\lambda} = \sum_{j=1}^M \langle \mathcal{P}_j(a_j(\mathbf{x}), \tau) \rangle \mathbf{v}_j = \tau \xi(\mathbf{x}),$$

where

$$\xi(\mathbf{x}) = \sum_{j=1}^M a_j(\mathbf{x}) \mathbf{v}_j \quad (56)$$

is the state change per unit time. Therefore, the average difference in the propensity function $a_j(\mathbf{x})$ is given by $|a_j(\mathbf{x} + \bar{\lambda}) - a_j(\mathbf{x})|$, which, according to the leap condition, should satisfies

$$|a_j(\mathbf{x} + \bar{\lambda}) - a_j(\mathbf{x})| \leq \varepsilon a_0(\mathbf{x}), \quad (j = 1, \dots, M), \quad (57)$$

where $0 < \varepsilon \ll 1$, and

$$a_0(\mathbf{x}) = \sum_{j=1}^M a_j(\mathbf{x}). \quad (58)$$

Since

$$|a_j(\mathbf{x} + \bar{\lambda}) - a_j(\mathbf{x})| \approx \bar{\lambda} \cdot \nabla a_j(\mathbf{x}) = \sum_{i=1}^N \tau \xi_i(\mathbf{x}) \frac{\partial a_j(\mathbf{x})}{\partial x_i}.$$

Let

$$b_{ji}(\mathbf{x}) = \frac{\partial a_j(\mathbf{x})}{\partial x_i}, \quad (59)$$

(57) can be approximated by

$$\tau \left| \sum_{j=1}^N \xi_i(\mathbf{x}) b_{ji}(\mathbf{x}) \right| \leq \varepsilon a_0(\mathbf{x}), \quad (j = 1, \dots, M),$$

which yields

$$\tau \leq \varepsilon a_0(\mathbf{x}) / \left| \sum_{j=1}^N \xi_i(\mathbf{x}) b_{ji}(\mathbf{x}) \right|.$$

Thus, the largest value of τ is given by

$$\tau = \min_{j \in [1, M]} \left\{ \varepsilon a_0(\mathbf{x}) / \left| \sum_{i=1}^N \xi_i(\mathbf{x}) b_{ji}(\mathbf{x}) \right| \right\}. \quad (60)$$

In [21] and [6], two successive refinements were made. The latest τ -selection procedure given in [6] is more accurate, easier to code, and faster to execute than the earlier procedures, but logically more complicated.

If the τ value generated above is much larger than the time required for the stochastic simulation algorithm, this approximate procedure will be faster than the exact stochastic simulation algorithm. However, if τ turns out to be less than a few multiples of the time required for the stochastic simulation algorithm to make an exact time step (in an order of $1/a_0(\mathbf{x})$), it would be better to use the stochastic simulation algorithm instead.

To avoid the negative populations in tau-leaping, several strategies have been proposed, in which the unbounded Poisson random numbers k_j are replaced by bonded binormal random numbers [7, 53]. In 2005, Cao et al. [3] proposed a new approach to resolve this difficulty. In this new Poisson tau-leaping procedure, the reaction channels are separated into two classes: *critical reactions* that may exhaust one of its reactants after some firings, and *noncritical reactions* otherwise. Next, the noncritical reactions are handled by the regular tau-leaping method to obtain a leap time τ' . And apply the exact stochastic simulation algorithm to the critical reactions, which gives the time τ'' and the index j_c of the next critical reaction. The actual time step τ is then taken to be the smaller

of τ' and τ'' . If the former, no critical reactions fires, and if the latter, only one critical reaction R_{j_c} fires.

Below are main steps for the modified tau-leaping procedure that can avoid negative populations (refer [3] for details)

1. **Initialization:** Let $\mathbf{X} = \mathbf{x}_0$, and $t = t_0$.
2. **Identify the critical reactions:** Identify the reaction channels R_j for which $a_j(\mathbf{X}) > 0$ and may exhaust one of its reactants after some firings.
3. **Calculate the leap time:** Compute the largest leap time τ' for the noncritical reactions.
4. **Monte Carlo step:** Generate (τ'', j_c) for the next critical reaction according to the modified density function according to (52).
5. **Determine next step:**
 - (a) If $\tau' < \tau''$: Take $\tau = \tau'$. For all critical reactions R_j , set $k_j = 0$. For all the noncritical reactions R_j , generate k_j as a Poisson random variable with mean $a_j(\mathbf{X})\tau$.
 - (b) If $\tau'' \leq \tau'$: Take $\tau = \tau''$. Set $k_{j_c} = 1$, and for all other critical reactions set $k_j = 0$. For all the noncritical reactions R_j , generate k_j as a Poisson random variable with mean $a_j(\mathbf{X})\tau$.
6. **Update:** Increase the time by τ , and replace the molecule count by $\mathbf{X} + \sum_{j=1}^M k_j \mathbf{v}_j$.
7. **Iterate:** Go back to Step 2 unless the simulation time has been exceeded.

5.3 Other simulation methods

In addition to the prominent approximate acceleration procedure tau-leaping, there are some other strategies that tend to speedup the stochastic simulation algorithm. Here we briefly outline two of the most promising methods.

Many real systems in biological processes involve chemical reactions with different time scales, “fast” reactions fire very much more frequently than “slow” ones. Procedures to handle such systems often involve a stochastic generalization of the quasi-steady-state assumption or partial (rapid) equilibrium methods of deterministic chemical kinetics [2, 5, 23, 43, 47]. The *slow-scale stochastic simulation algorithm* (ssSSA) (or *multiscale stochastic simulation algorithm*) is a systematic procedure for partitioning the system into fast and slow reactions, and only simulate the slow reactions by specially modified propensity functions [2, 4, 5].

Another approach to simulate multiscale chemical reaction systems include different kinds of *Hybrid methods* [1, 13, 24, 46]. Hybrid methods combine the deterministic reaction rate equation with the stochastic simulation algorithm. The idea is to split the system into two regimes: the continuous regime of large molecule population species, and the discrete regime of small molecule population species. The continuous regimes is treated by ordinary differential equations, while the discrete regime is simulated by the stochastic simulation algorithm. Hybrid methods efficiently utilize the multiscale properties of the

system. However, because of lacking a rigorous theoretical foundation, there are still many unsolved problems [33].

There are many approaches trying to find the numerical solution of the chemical master equation [9, 11, 12, 25, 26, 28, 30, 37]. In addition, numerical methods for the Langevin equation have been well documented (refer [31] for example). We will not get into these two subjects here.

6 Summary

In modeling a well-stirred chemical reacting systems, the chemical master equation provides an ‘exact’ description of the time evolution of the states. However, it is difficult to directly study the chemical master equation because of the dimension problem. Several approximations are therefore developed, including the Fokker-Plank equation, reaction rate equation, and chemical Langevin equation. The reaction rate equation is widely used when fluctuations are not important. When the fluctuations are significant, the chemical Langevin equation can provide reasonable description for the statistical properties of the kinetics, despite the conditions in deriving the chemical Langevin equation may not hold.

When there are noise perturbations to the kinetic parameters, there is no simple way to model the system dynamics because the time dependence of environment variables can be very complicated. In a particular case, we can replace the reaction rates by log-normal random variables, and generalize the reaction rate equation or the chemical Langevin equation to describe the dynamics of a chemical system with extrinsic noise.

Stochastic simulation algorithm (SSA) is an ‘exact’ numerical simulation that shares the same fundamental basis as the chemical master equation. The approximate explicit tau-leaping produce, on the other hand, is closely relate to the chemical Langevin equation. The robustness and efficiency of the two methods have been considerable improved in recent years, and these procedures seems to be nearing maturity [20]. In the last few years, some other strategies have been developed for simulating the systems that are dynamically stiff [20, 33].

We conclude this paper with Figure 4, which summarizes the theoretical structure of stochastic modeling for chemical kinetics (also refer [20]).

References

- [1] A. Alfonsi, E. Cancés, G. Turinici, B. D. Ventura, and W. Huisings. Exact simulation of hybrid stochastic and deterministic models for biochemical systems. Technical Report RR-5435, INRIA, 2004.
- [2] Y. Cao, D. T. Gillespie, and L. R. Petzold. Accelerated stochastic simulation of the stiff enzyme-substrate reaction. *J. Chem. Phys.*, 123:144917, 2005.
- [3] Y. Cao, D. T. Gillespie, and L. R. Petzold. Avoiding negative populations in explicit poisson tau-leaping. *J. Chem. Phys.*, 123:054104, 2005.

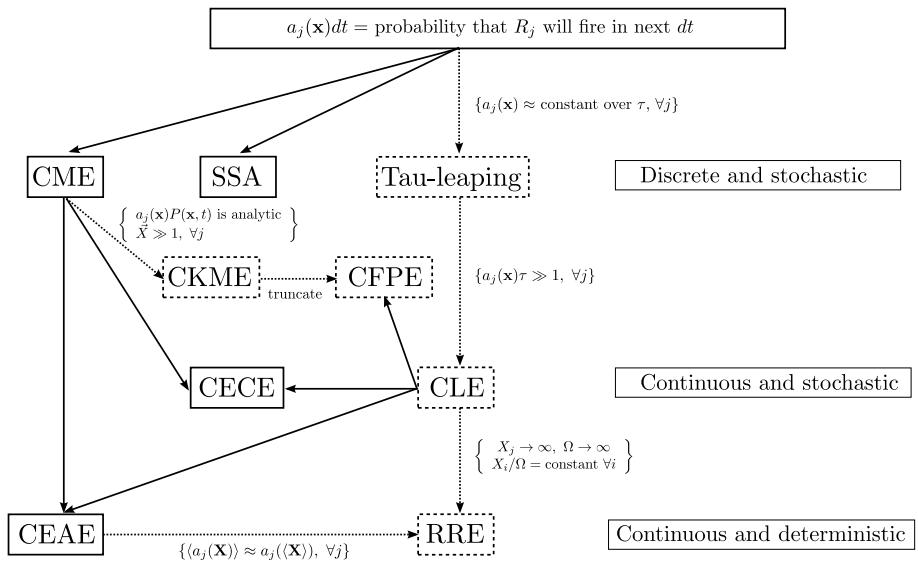


Figure 4: (Modified from ref. [20]) Theoretical structure of stochastic chemical kinetics. Everything originate from the fundamental premise of propensity function given at the top box. Solid arrows show the exact inferences routes, and dotted arrows show the approximate ones, with justified conditions in braces. Solid boxes are exact results: the chemical master equation (CME), the stochastic simulation algorithm (SSA), the chemical ensemble coefficient equation (CECE), and the chemical ensemble average equation (CEAE). Dashed boxes are approximate results: the Tau-leaping algorithm, the chemical Langevin equation (CLE), the chemical Kramers-Moyal equation (CKME), the chemical Fokker-Plank equation (CFPE), and the reaction rate equation (RRE).

- [4] Y. Cao, D. T. Gillespie, and L. R. Petzold. Multiscale stochastic simulation algorithm with stochastic partial equilibrium assumption for chemically reacting systems. *J. Comput. Phys.*, 206:395–411, 2005.
- [5] Y. Cao, D. T. Gillespie, and L. R. Petzold. The slow-scale stochastic simulation algorithm. *J. Chem. Phys.*, 122:014116, 2005.
- [6] Y. Cao, D. T. Gillespie, and L. R. Petzold. Efficient step size selection for the tau-leaping simulation method. *J. Chem. Phys.*, 124:044109, 2006.
- [7] A. Chatterjee, D. Vlachos, and M. Katsoulakis. Binomial distribution based τ -leap accelerated stochastic simulation. *J. Chem. Phys.*, 122:024112, 2005.
- [8] E. L. Crow and K. Shimizu, editors. *Lognormal Distributions: Theory and Applications*. New York, 1988.
- [9] P. Deuflhard, W. Huisenga, T. Jahnke, and M. Wulkow. Adaptive discrete galerkin methods applied to the chemical master equation. *SIAM J. Sci. Comput.*, 30:2990–3011, 2008.
- [10] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain. Stochastic gene expression in a single cell. *Science*, 297:1183–1186, 2002.
- [11] S. Engblom. Galerkin spectral method applied to the chemical master equation. *Commun. Comput. Phys.*, 5:871–896, 2009.
- [12] S. Engblom. Spectral approximation of solutions to the chemical master equation. *J. Comput. Appl. Math.*, 229:208–221, 2009.
- [13] R. Erban, I. G. Kevrekidis, D. Adalsteinsson, and T. C. Elston. Gene regulatory networks: a coarse-grained, equation-free approach to multiscale computation. *J. Chem. Phys.*, 124:084106, 2006.
- [14] S. Fiering, E. Whitelaw, and D. Martin. To be or not to be active: the stochastic nature of enhancer action. *Bioessays*, 22:381–387, 2000.
- [15] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.*, 22:403–434, 1976.
- [16] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81:2340–2361, 1977.
- [17] D. T. Gillespie. *Markov Processes: An Introduction for Physical Scientists*. Academic Press, San Diego, CA, 1992.
- [18] D. T. Gillespie. The chemical langevin equation. *J. Chem. Phys.*, 113:297–306, 2000.
- [19] D. T. Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *J. Chem. Phys.*, 115:1716–1733, 2001.
- [20] D. T. Gillespie. Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.*, 58:35–55, 2007.

- [21] D. T. Gillespie and L. R. Petzold. Improved leap-size selection for accelerated stochastic simulation. *J. Chem. Phys.*, 119:8229–8234, 2003.
- [22] I. Golding, J. Paulsson, S. M. Zawilski, and E. C. Cox. Real-time kinetics of gene activity in individual bacteria. *Cell*, 123:1025–1036, 2005.
- [23] J. Goutsias. Quasiequilibrium approximation of fast reaction kinetics in stochastic biochemical systems. *J. Chem. Phys.*, 122:184102, 2005.
- [24] E. L. Haseltine and J. B. Rawlings. Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *J. Chem. Phys.*, 117:6959–6969, 2002.
- [25] M. Hegland, A. Hellander, and P. Lötstedt. Sparse grids and hybrid methods for the chemical master equation. *BIT*, 48:265–283, 2008.
- [26] T. Jahnke. An adaptive wavelet method for the chemical master equation. *SIAM J. Sci. Comput.*, 31:4373–4394, 2010.
- [27] T. Jahnke and W. Huisenga. Solving the chemical master equation for monomolecular reaction systems analytically. *J. Math. Biol.*, 54:1–26, 2007.
- [28] T. Jahnke and W. Huisenga. A dynamical low-rank approach to the chemical master equation. *Bull. Math. Biol.*, 70:2283–2302, 2008.
- [29] M. Kærn, T. C. Elston, W. J. Blake, and J. J. Collins. Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.*, 6:451–464, 2005.
- [30] C. F. Khoo and M. Hegland. The total quasi-steady state assumption: its justification by singular perturbation and its application to the chemical master equation. *ANZIAM J.*, 50:C429–C443, 2008.
- [31] P. E. Kloeden and E. Platen. *Numerical solutions of stochastic differential equation*. Springer-Verlag, New York, 1992.
- [32] J. Lei. Stochasticity in single gene expression with both intrinsic noise and fluctuation in kinetic parameters. *J. Theor. Biol.*, 256:485–492, 2009.
- [33] H. Li, Y. Cao, L. R. Petzold, and D. T. Gillespie. Algorithms and software for stochastic simulation of biochemical reacting systems. *Biotech. Prog.*, 24:56–61, 2008.
- [34] E. Limpert, W. A. Stahel, and M. Abbt. Log-normal distributions across the sciences: keys and clues. *BioScience*, 51:341–352, 2001.
- [35] H. H. McAdams and A. P. Arkin. Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. USA*, 94:814–819, 1997.
- [36] H. H. McAdams and A. P. Arkin. It’s a noisy business! genetic regulation at the nanomolar scale. *Trends Genet.*, 15:65–69, 1999.
- [37] B. Munsky and M. Khammash. A multiple time interval finite state projection algorithm for the solution to the chemical master equation. *J. Comput. Phys.*, 226:818–835, 2007.

- [38] B. Øksendal. *Stochastic Differential Equations*. Springer, Berlin, 6 edition, 2005.
- [39] I. Oppenheim, K. E. Shuler, and G. H. Weiss. Stochastic and deterministic formulation of chemical rate equations. *J. Chem. Phys.*, 50:460–466, 1969.
- [40] J. Paulsson. Summing up the noise in gene networks. *Nature*, 427:415–418, 2004.
- [41] J. M. Pedraza and A. van Oudenaarden. Noise propagation in gene networks. *Science*, 307:1965–1969, 2005.
- [42] A. Raj and A. van Oudenaarden. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*, 135:216–226, 2008.
- [43] C. V. Rao and A. P. Arkin. Stochastic chemical kinetics and the quasi-steady-state assumption: application to the gillespie algorithm. *J. Chem. Phys.*, 118:4999–5010, 2003.
- [44] C. V. Rao, D. M. Wolf, and A. P. Arkin. Control, exploitation and tolerance of intracellular noise. *Nature*, 420:231–237, 2002.
- [45] N. Rosenfeld, J. W. Young, U. Alon, P. S. Swain, and M. B. Elowitz. Gene regulation at the single-cell level. *Science*, 307:1962–1965, 2005.
- [46] H. Salis and Y. Kaznessis. Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions. *J. Chem. Phys.*, 122:054103, 2005.
- [47] A. Samant and D. G. Vlachos. Overcoming stiffness in stochastic simulation stemming from partial equilibrium: a multiscale monte carlo algorithm. *J. Chem. Phys.*, 123:144114, 2005.
- [48] M. S. Samoilov, G. Price, and A. P. Arkin. From fluctuations to phenotypes: the physiology of noise. *Sci STKE*, 2006:re17, 2006.
- [49] R. Schlicht and G. Winkler. A delay stochastic process with applications in molecular biology. *J. Math. Biol.*, 57:613–648, 2008.
- [50] V. Shahrezaei and P. S. Swain. Analytical distributions for stochastic gene expression. *Proc. Natl. Acad. Sci. USA*, 105:17256–17261, 2008.
- [51] V. Shahrezaei and P. S. Swain. The stochastic nature of biochemical networks. *Curr. Opin. Biotechnol.*, 19:369–374, 2008.
- [52] S. S. Sommer and N. Adam Rin. The lognormal distribution fits the decay profile. *Biochem. Biophys. Res. Commu.*, 90(1):135–141, 1979.
- [53] T. Tian and K. Burrage. Binomial leap methods for simulating stochastic chemical kinetics. *J. Chem. Phys.*, 121:10356–10364, 2004.
- [54] G. E. Uhlenbeck and L. S. Ornstein. On the theory of the brownian motion. *Physical Review*, 36:823–841, 1930.
- [55] N.G. van Kampen. Langevin-like equation with colored noise. *J. Statis. Phys.*, 54(5/6):1289–1308, 1989.

- [56] N.G. van Kampen. *Stochastic Processes in Physics and Chemistry*. North-Holland, Amsterdam, 3rd edition, 2007.
- [57] J. Vilar, H. Y. Kueh, N. Barkai, and S. Leibler. Mechanisms of noise-resistance in genetic oscillators. *Proc. Natl. Acad. Sci. USA*, 99:5988–5992, 2002.